Using singleton densities to detect recent selection in Bos taurus

doi:10.1002/evl3.263

Using singleton densities to detect recent selection in *Bos taurus*

Matthew Hartfield,[1,2,3] Nina Aagaard Poulsen,[4] Bernt Guldbrandtsen,[5,6,7] and Thomas Bataillon[1]

[1] *Bioinformatics Research Centre, Aarhus University, Aarhus DK-8000, Denmark*

[2] *Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3FL, United Kingdom*

[3] *E-mail: m.hartfield@ed.ac.uk*

[4] *Department of Food Science, Aarhus University, Aarhus N DK-8200, Denmark*

[5] *Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Tjele DK-8830, Denmark*

[6] *Rheinische Friedrich-Wilhelms-Universität Bonn, Institut für Tierwissenschaften, Bonn DE-53115, Germany*

[7] *Department of Veterinary Sciences, Copenhagen University, Frederiksberg C DK-1870, Denmark*

Received June 2, 2020
Accepted October 8, 2021

Many quantitative traits are subject to polygenic selection, where several genomic regions undergo small, simultaneous changes in allele frequency that collectively alter a phenotype. The widespread availability of genome data, along with novel statistical techniques, has made it easier to detect these changes. We apply one such method, the "Singleton Density Score" (SDS), to the Holstein breed of *Bos taurus* to detect recent selection (arising up to around 740 years ago). We identify several genes as candidates for targets of recent selection, including some relating to cell regulation, catabolic processes, neural-cell adhesion and immunity. We do not find strong evidence that three traits that are important to humans—milk protein content, milk fat content, and stature—have been subject to directional selection. Simulations demonstrate that because *B. taurus* recently experienced a population bottleneck, singletons are depleted so the power of SDS methods is reduced. These results inform on which genes underlie recent genetic change in *B. taurus*, while providing information on how polygenic selection can be best investigated in future studies.

KEY WORDS: *Bos taurus*, genomics, milk fat, milk protein, selection, stature.

Impact Statement

Many traits of ecological or economic importance (including height, disease propensity, climatic adaptation) are "polygenic." That is, they are affected by a large number of genetic variants, with each one only making a small contribution to a trait, but collectively influence variation. As selection acts on all of these variants simultaneously, it only changes the frequency of each one by a small amount, making it hard to detect such selection from genome data. This situation has changed in recent years, with the proliferation of whole-genome data from many individuals, along with the development of methods to detect the subtle effects of polygenic selection. Here, we use data from 102 genomes from domesticated cattle (*Bos taurus*) that has experienced intense artificial selection since domestication, and test whether we can detect signatures of recent selection (arising up to 740 years ago). Domesticated species are appealing for this kind of study, as they are subject to extensive genome sequencing studies, and genetic variants can be related to traits under selection. We carried out our analysis in two parts. We first performed a genome-wide scan to find individual genetic regions that show signatures of recent selection. We identify those relating to cell regulation, catabolic processes, neural-cell adhesion, and immunity. In the second part, we then analyzed genetic regions associated with three key traits: milk protein content, milk fat content, and stature. We tested whether these regions collectively

showed a signature of selection, but did not find a significant result in any of these cases. Simulations suggest that the domestication history of cattle affected the power of these methods. We end with a discussion on how to best detect polygenic selection in future studies.

Determining which genomic regions have been subject to selection is a major research goal in evolutionary genetics. Traditional methods have focused on detecting strong selection affecting individual genes (Nielsen, 2005; Vitti et al., 2013; Stephan, 2019). An alternative process is "polygenic selection," where many loci contribute to genetic variation in a trait, so selection acting on it is expected to generate small and simultaneous allele frequency changes at multiple loci (Pritchard & Di Rienzo, 2010; Pritchard et al., 2010). Many polygenic models have been formulated to account for both the response to phenotypic selection and the maintenance of genetic variance in quantitative traits (reviewed by Sella and Barton [2019] and Barghi et al. [2020]). Among them is Fisher's infinitesimal model, which is important for its historical role in uniting population and quantitative genetics, and its recent renaissance in the context of genome-wide association studies (Fisher, 1918; Barton & Keightley, 2002; Barton et al., 2017; Charlesworth & Edwards, 2018; Visscher & Goddard, 2019). Although it has been possible to identify which genetic regions contribute to trait variation, it has historically been hard to infer which alleles have been involved in the polygenic selection response. Extensive theoretical studies of how alleles at multiple loci act when a population adapts to a new optimum generally find that "large-effect" alleles, which strongly affect a trait, are the first to spread and fix, whereas "small-effect" alleles take much longer to reach high frequencies (de Vladar and Barton, 2014; Wollstein & Stephan, 2014; Jain & Stephan, 2015, 2017a, 2017b; Stetter et al., 2018; Thornton, 2019; Hayward & Sella, 2019). Furthermore, if epistasis exists between variants, many selected alleles do not reach fixation as they eventually become deleterious (de Vladar & Barton, 2014; Jain & Stephan, 2017b). The spread of large-effect alleles may also be impeded if a faster adaptive response can be otherwise realized through changes at many small-effect alleles (Lande, 1983; Chevin & Hospital, 2008; Pavlidis et al., 2012; Chevin, 2019). Alternatively, if the optimum shift is sufficiently big, then large-effect mutations that first go to fixation can subsequently be replaced by small-effect variants over longer timescales (on the order of the population size; Hayward and Sella [2019]). Overall, only a small proportion of loci affected by polygenic selection are expected to fix sufficiently quickly to leave selection signatures in genomic data (Pavlidis et al., 2012; Thornton, 2019).

Due to this difficulty, earlier methods for detecting polygenic selection focused on cases where selection favors distinct phenotypes in different populations, so trait differentiation among populations will be greater than expected under neutral drift. Tests for this form of selection relied on comparing Q_{ST} and F_{ST} statistics, which, respectively, measured mean genetic differentiation at the trait itself and a set of neutral loci (Whitlock, 2008; Le Corre & Kremer, 2012; Savolainen et al., 2013). Yet these methods do not determine which genomic regions are subject to selection. This situation has now changed with the increased number of genome-wide association study (GWAS) data that link genotypes and phenotypes, as exemplified by the development of large cohort studies (e.g., the UK Biobank; Bycroft et al. 2018). The release of these data spurred a series of studies and new methods designed specifically to detect polygenic selection. These methods usually involve determining which SNPs affecting a phenotype show correlated changes in frequency (Berg & Coop, 2014; Racimo et al., 2018; Sanjak et al., 2018; Josephs et al., 2019; Berg et al., 2019a, 2019b; Uricchio et al., 2019; Edge & Coop, 2019; Kreiner et al., 2020; Wieters et al., 2021; Gramlich et al., 2021); which sets of alleles are associated with certain environmental or climatic variations (Coop et al., 2010; Turchin et al., 2012; Robinson et al., 2015; Yeaman et al., 2016; Exposito-Alonso et al., 2018; Zan & Carlborg, 2018; Exposito-Alonso et al., 2019; MacLachlan et al., 2021; Ehrlich et al., 2021; Fuhrmann et al., 2021; Rowan et al., 2021); or determining which SNPs or genetic regions explain a large fraction of phenotypic variance and trait heritability (Zhou et al., 2013; Yang et al., 2015; Gazal et al., 2017; Zeng et al., 2018; Schoech et al., 2019; Exposito-Alonso et al., 2020; Duntsch et al., 2020; Zeng et al., 2021). Some of these approaches use overlapping methods.

Detecting recent polygenic selection is much harder, as long periods of time (number of generations on the order of the population size; Hayward & Sella, 2019; Thornton, 2019) may be needed to cause detectable frequency changes in alleles with small effect sizes. Over shorter timescales, these frequency changes are expected to be more modest and harder to detect (Stephan, 2016; Jain & Stephan, 2017a). A recent breakthrough in detecting these subtle changes was the development of the "Singleton Density Score" (SDS), a statistic tailored to detect recent and coordinated allele frequency changes over many SNPs (Field et al., 2016). Recent selection at a locus favoring one variant will lead to a reduction in the number of singletons (i.e., variants that are only observed once) around it. The SDS detects regions that exhibit a reduction in the density of singletons, to determine candidate regions that have been subject to recent selection. Using this approach, Field et al. (2016) found correlations between SDS scores at SNPs and their associated GWAS effect sizes for several polygenic traits in the modern UK human population, including increased height, infant head circumference, and fasting insulin. Their findings suggested that these traits have been subject to recent selection during the most recent 75 or so

generations (about 2000 years). However, these (and other) results that detect selection for increased height may instead reflect previously unaccounted-for population structure (Novembre & Barton, 2018; Barton et al., 2019; Sohail et al., 2019; Berg et al. 2019a; Uricchio et al., 2019; Edge & Coop, 2019).

The SDS method is ideally suited to organisms where large amount of whole-genome data are available, along with quantitative trait loci (QTL) or GWAS information that link genotypes to phenotypes. Domesticated species are attractive systems for studying recent selection, as selected phenotypes are often already known and they are subject to large-scale sequencing studies. Investigating the genetic architecture underlying rapid selection in these species is also important to determine how they respond to agricultural practices, and uncover selection targets that can be used to improve breeding programs (Georges et al., 2019). Domestic cattle *Bos taurus* has been subject to intensive genomics analyses to improve artificial selection for traits that are important for human use, including milk protein yield, milk fat content, and stature (Hayes et al., 2009; Meuwissen et al., 2013; Wray et al., 2019). These traits are influenced in part by an individual's genome, with significant heritability estimates being recorded, some as high as 80% (Soyeurt et al., 2007; Haile-Mariam et al., 2013; Buitenhuis et al., 2016). Previous selection scans on *B. taurus* reported individual regions that were likely to be subject to recent selection, some of which were close to genetic regions for stature, milk protein content, and milk fat content (Lemay et al., 2009; MacEachern et al., 2009; Qanbari et al., 2010; Boitard & Rocha, 2013; Qanbari et al., 2014; Zhao et al., 2015; Boitard et al., 2016a; Bouwman et al., 2018). However, these traits are polygenic, with several genetic regions and QTLs associated with each (Lemay et al., 2009; Boitard et al., 2016a; Bouwman et al., 2018; van den Berg et al., 2020). Although recent methods have been developed to detect polygenic environmental adaptation (Rowan et al., 2021), there has yet to be a formal test of whether these intrinsic traits show evidence of polygenic selection.

Here, we applied the SDS method to whole-autosome sequencing data from 102 *B. taurus* Holstein individuals. We first determined genetic regions that have been subject to recent directional selection, and subsequently tested whether evidence exists for recent selection acting on a set of QTLs underlying either milk protein content, milk fat content, or stature in this breed.

Results

METHODS OUTLINE

We filtered the data to retain only biallelic SNPs that had a sensible level of coverage and did not lie in putatively over-assembled regions (i.e., duplicated sections that caused many reads to assemble at a specific genetic location). Over-assembled regions appear as highly heterozygous with elevated coverage, and can exhibit false signatures of recent selection. We also obtained a set of singletons and filtered them to retain high-quality variants where both alleles were equally well covered to remove potentially erroneous calls. We polarized test SNPs using outgroup sequences and applied the SDS test of Field et al. (2016) to detect recent selection, with increased SDS values reflecting selection favoring derived SNPs over ancestral variants. We standardized SDS scores with those of a similar frequency, so they are normally distributed (similar normalization was also carried out by Field et al. 2016). These values are denoted sSDS for "standardized SDS." Further details are available in the *Methods* in the Supporting Information.

ESTIMATING TIMESCALE OF SELECTION

We first determined the timescale over which we expect to detect selection in *B. taurus* using the SDS method. SDS measures the changes in singleton numbers around putatively selected SNPs, relative to background numbers in the absence of selection. As singletons arise on the tips of the underlying gene trees, the average tip length in the genealogy of sequenced samples determines the timescale over which the SDS detects a signal (Field et al., 2016). As more haploid genomes are included in the study, the time to first coalescence between two samples decreases, reducing the tip lengths and therefore shortening the timescale over which SDS detects selection (Field et al., 2016). We hence simulate tip-ages over a range of sample sizes to investigate how this timescale changes accordingly.

To calculate the mean tip age, we simulated gene genealogies under two scenarios. We first simulated the Holstein population demography inferred by Boitard et al. (2016b), which suggested that this population experienced a sudden decline in effective population size (N_e) since domestication, but with a present-day N_e ($\sim$793) that is much larger than that inferred from pedigree data ($\sim$49; Sørensen et al. 2005) or from temporal variation in SNP frequencies ($\sim$48; Jiménez–Mena et al., 2016). Hence, we also simulated genealogies under a second model that used the Boitard et al. (2016b) demographic model, but with the present-day N_e set to 49. These scenarios will be referred to as the "High N_0" and "Low N_0" models, respectively.

Figure 1 shows simulation results. Depending on the assumed present-day N_e, the tip length in our sample of 204 alleles (i.e., assuming two per diploid individual) goes back either 65 or 148 generations. Assuming 5 years per generation (Boitard et al., 2016b), this timescale corresponds to between 325 and 740 years ago. Since *B. taurus* domestication started around 10,000 years ago (Zeder, 2008), the sample size used in this study will only capture selection acting in the very recent past that is more relevant for breed formation, rather than selection during *B. taurus* domestication. Sample sizes and tip-ages are linearly related on a

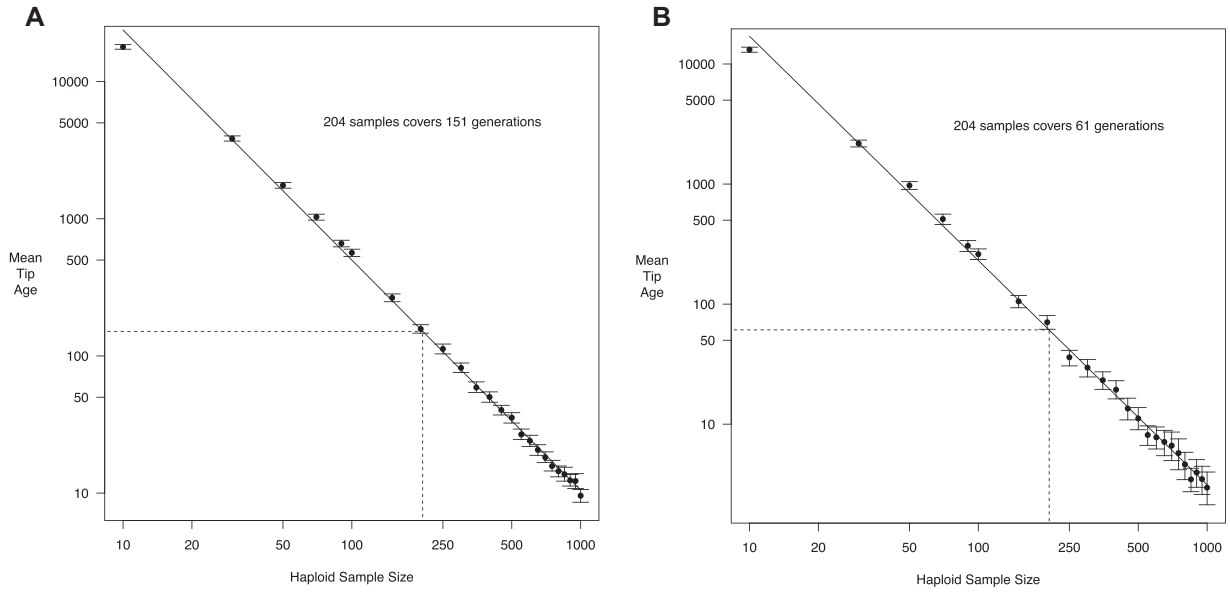

Figure 1. Simulated mean tip age for *B. taurus*, as a function of the number of haploid samples. Simulations assumed either (A) demography as inferred by Boitard et al. (2016b) (the "High N_0" model) or (B) the same but with a smaller present-day N_e of 49 (the "Low N_0" model). Points are the mean values; bars show 95% confidence intervals. The solid line is the best linear fit to the log of both values; dotted lines show the predicted tip age for 204 alleles.

Table 1. Genes that overlap or lie close to Bonferroni-significant sSDS regions. The "High, Low N_0" column specifies which genes are close to significant SNPs for each N_0 model.

Chromosome	Gene Name	Start Position	End Position	Gene Biotype	High, Low N_0
1	PPM1L	106405113	106727070	Protein Coding	High, Low
5	TMCC3	24306913	24595494	Protein Coding	High, Low
5	CEP83	24070404	24345243	Protein Coding	High, Low
17	U6	43381106	43381209	snRNA	Low
17	CTSO	43364999	43381605	Protein Coding	Low
17	TDO2	43386894	43403747	Protein Coding	High, Low
23	OR12D2H	29291787	29292713	Protein Coding	High, Low
23	OR12D2E	29305933	29309785	Protein Coding	High, Low
24	GAREM1	24694637	24927333	Protein Coding	High, Low
29	NTM	34576918	34994005	Protein Coding	High, Low

log-log scale, meaning that an increase in sample size will greatly decrease the timescale over which SDS detects selection. For example, with 500 haplotypes then SDS will detect selection acting no more than 50 generations ago, depending on the underlying demographic model.

We will focus on detecting selection signatures assuming the high N_0 model. Results using the low N_0 model to calibrate scores were broadly similar. They are outlined in the Supporting Information; we will highlight when differences arise.

GENOME-WIDE sSDS

Figure 2 plots sSDS values (at SNPs with minor allele frequency greater than 5%) across all autosomes, excluding chromosome 25 (due to an insufficient number of singletons needed to obtain SDS scores after filtering). Many SNPs have elevated sSDS scores (158 SNPs at FDR < 0.05; 306 for the low N_0 model). Several regions contain SNPs with significantly high sSDS values (Bonferroni-corrected nominal $P < 0.05$; actual $P < \sim 2.7 \times 10^{-8}$). To further investigate potential selection targets, we looked for genes that either overlapped significant SNPs or lay 10-kb up- or downstream of them. Linkage disequilibrium (LD), as measured by r^2, decays to around 0.2 over 50 kb in Danish Holstein breeds (Buitenhuis et al., 2016), so genes within 10 kb should be in LD with regions harboring high sSDS scores. Table 1 lists these genes, with more targets present under the low N_0 model. Most of these genes are of unknown function (as listed on UniProt); the list also includes an snRNA. *PPM1L* is involved with cellular regulation and the activation of stress-activated

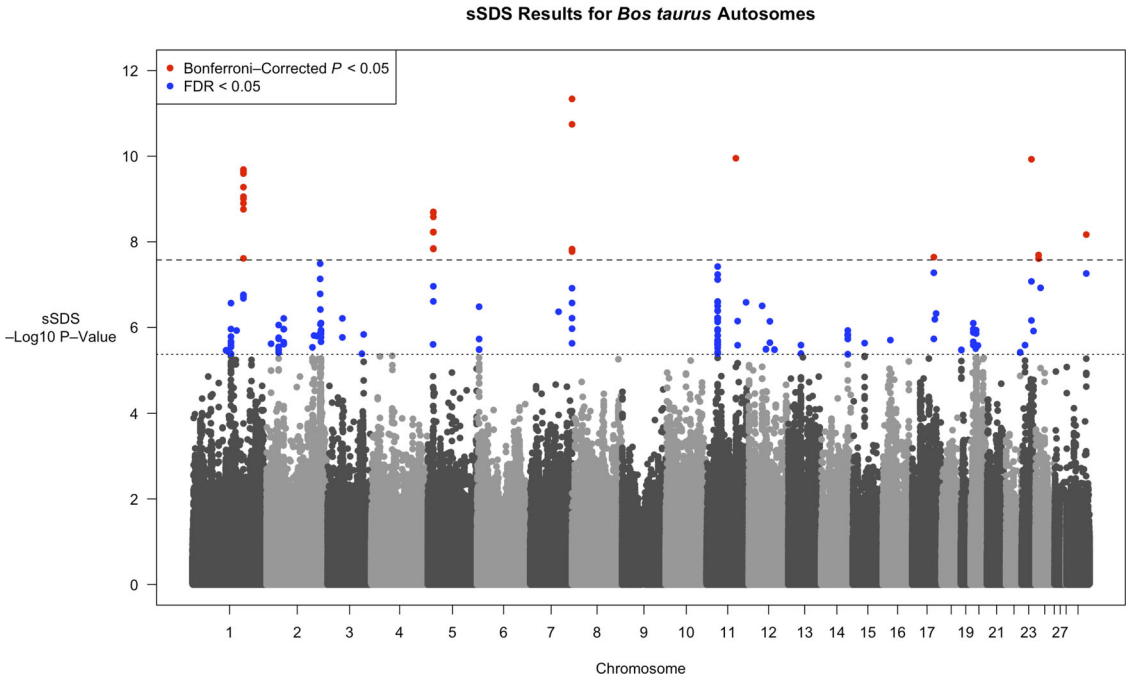

Figure 2. ***P*-values of sSDS scores across *B. taurus* autosomes, as plotted on a negative Log_{10} scale, as a function of the chromosome. Alternating black and gray points show (nonsignificant) values from different chromosomes. Blue points are SNPs with FDR < 0.05, with the cutoff denoted by a horizontal dotted line. Red points are SNPs with Bonferroni-corrected *P*-value < 0.05 (actual *P*-value < ~2.7 x 10^{-8}), with the cutoff denoted by a horizontal dashed line. Figure S1 shows results for the Low N_0 model.**

protein kinases. *TDO2* is involved in tryptophan-related catabolic processes, whereas *NTM* is implicated in neural cell adhesion. SNPs with significantly elevated scores are also found on chromosome 23 near the MHC region, which may reflect overdominant selection. All Bonferroni-significant SNPs were removed from subsequent tests of recent polygenic selection to prevent directional selection from skewing the underlying sSDS distributions. Figure S1 shows results for the low N_0 model.

TESTING FOR POLYGENIC SELECTION ACTING ON MILK FAT, PROTEIN AND STATURE

If polygenic selection were acting on specific traits, we expect a positive correlation between the effect sizes of variants underpinning a trait, and selection acting on them as measured by sSDS. We collated sSDS scores of SNPs that lie close to QTLs reported for either milk fat percentage, milk protein percentage (van den Berg et al., 2020), or those that lie close to stature QTLs (Bouwman et al., 2018). The latter were inferred from a meta-analysis of GWAS studies conducted in seven Holstein populations, but not every QTL had an effect size reported in each population. We hence investigated two overlapping consensus QTL sets, where an effect size was either reported in at least six of seven populations (yielding 42 QTLs with sSDS scores associated with them) or where effect sizes were reported in at least five of seven populations (58 QTLs had sSDS scores). We repolarized SDS scores so that a positive score reflected a trait-increasing effect; we denote these values "tSDS" following Field et al. (2016). We then determine if there was a positive correlation between the absolute log_{10}-value of the QTL *P*-value (a proxy for the effect size) and tSDS.

Figure 3 shows the relationship between QTL *P*-values and tSDS for SNPs that lie close to QTLs. Although positive trends are observed as determined using a linear model, they all exhibit nonsignificant correlations (milk fat percentage Spearman $\rho = 0.0990$, $P = 0.603$; milk protein percentage Spearman $\rho = 0.0354$, $P = 0.758$; stature from six breeds Spearman $\rho = -0.0739$, $P = 0.642$; stature from five breeds Spearman $\rho = -0.00966$, $P = 0.943$). Relationships remain nonsignificant after removing an outlier point for the milk traits whose QTL has an extremely low *P*-value (Figure S2), and also under the low N_0 model (Figure S3; see figure legends for correlation *P*-values).

sSDS (and tSDS) can become correlated along the genome if focal SNPs are in LD with one another, which was not accounted for in the preceding analyses. To determine whether LD could have affected these correlations, we randomly subsampled sSDS scores from SNPs that shared the same chromosome and bin of derived-allele frequency as the SNPs used in the above analyses, and repolarized them to transform them into tSDS

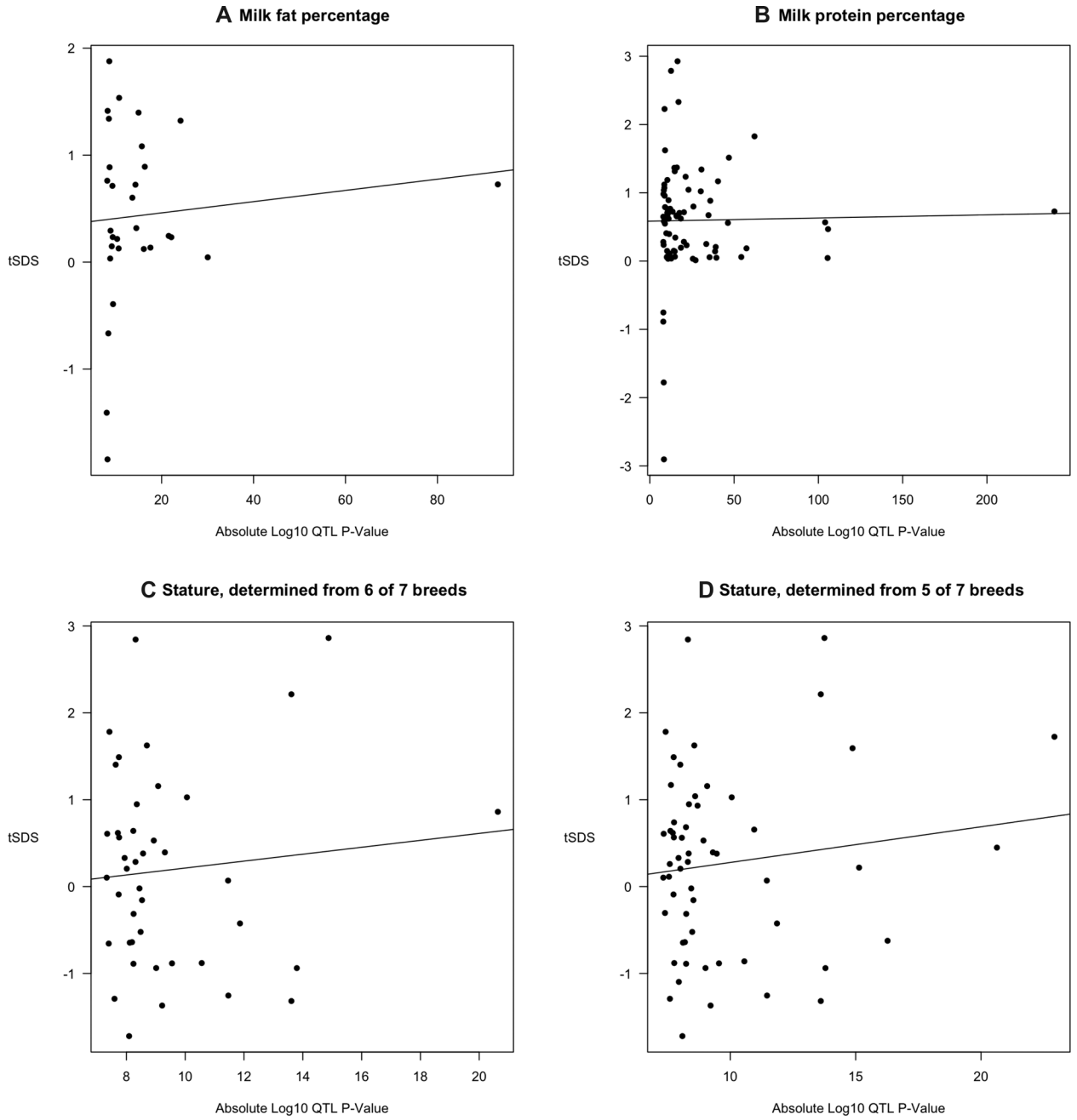

Figure 3. Relationship between tSDS scores near milk or stature QTLs, as noted in the subheadings, and the absolute log *P*-value of QTLs. Lines show a linear model regression fit. Figure S3 shows results assuming a low N_0 model.

values. We then determined the Spearman's ρ associated with these permuted values to determine whether that for the true data was significantly elevated (see *Methods* for details). In all cases, the observed value was not significantly higher than for permuted values (see Figure S4 for histograms and exact *P*-values, which all exceed 0.05). We therefore conclude that these QTL datasets do not harbor SNPs with significantly different tSDS scores compared to the rest of the genome.

Discussion

SUMMARY OF RESULTS

We analyzed an extensive *B. taurus* genomic dataset to identify signatures of recent selection in the Holstein breed, and to determine whether the data contained a signal of polygenic selection acting on milk proteins and QTLs underlying phenotypic variation in stature. Given the sample size and the demographic history of Holsteins, the SDS method can detect very recent

selection events arising no more than approximately 740 years ago (Figure 1). A whole-genome scan for sSDS scores identified several targets of recent directional selection that overlap or lie close to protein-coding genes (Figure 2; Table 1). The genes whose functions are known are involved in protein regulation, catabolic processes, and neural-cell adhesion. Significant values were also observed near the MHC region. We subsequently investigated whether either milk protein genes or SNPs near stature QTLs collectively showed evidence of polygenic selection. We did so by testing whether there is a relationship between the QTL effect size, as measured by its *P*-value, and tSDS values to SNPs near them. However, no relationship was observed, even after performing a permutation test (Figs. 3 and S2-S4). Hence, although sSDS could reveal specific instances of recent selection, tests based on collective scores of variants associated with known selected traits yielded no signal of polygenic selection.

POTENTIAL REASONS FOR A LACK OF POLYGENIC SELECTION SIGNAL

Impact of Holstein demographic history

Although the SDS method detected individual candidate genes for very recent selection, we were unable to find strong evidence for polygenic selection acting on three traits that were subject to artificial selection since domestication. This result is *a priori* surprising, given that these traits have been subject to recent intense artificial selection. Recent studies generally find nonzero heritability estimates for them, indicating that there should be the potential for genetic variants underpinning them to change in response to artificial selection (Soyeurt et al., 2007; Haile-Mariam et al., 2013; Buitenhuis et al., 2016). In addition, the ratio of the mutation and recombination rates in cattle is just over three (Boitard et al., 2016b; Harland et al., 2018), indicating that several informative SNPs exist per haplotype that should improve the power of the SDS method (in contrast, this ratio is approximately equal to one in humans; Field et al., 2016).

One potential reason for this lack of signal is due to the population history of *Bos taurus*. The effective population size of many *B. taurus* breeds appears to have undergone a decline since domestication (Sørensen et al., 2005; Boitard et al., 2016b), which likely reflects successive bottlenecks due to domestication, breed formation, and intense recent selection. Population size reductions are known to reduce the number of low-frequency variants and increase the prevalence of intermediate-frequency variants (Harpending et al., 1998), which can affect the power of the SDS method. To understand if the history of *B. taurus* affects the detection of recent selection in Holstein cattle using SDS, we ran coalescent simulations to determine its ability to detect ongoing selection, given realistic Holstein population history and genetic parameters (see *Methods* for details). We simulated a partial sweep occurring in the middle of a 10-Mb region, either assuming a mutation rate in line with what has been inferred for Holstein or one 10-fold higher to replicate diversity expected in a genetic region with an elevated mutation rate.

For the standard mutation rate, no SDS scores were produced for any simulations. After inspecting the simulation results, we see that there is a large skew in the distribution of singleton numbers per individual with a large number of individuals (over 20 on average) that do not carry singletons at the end of simulations, preventing the calculation of a local SDS score (Figure 4). This fraction remained the same irrespective of whether the simulated region was neutral or subject to selection; the main effect of a sweep was to reduce the mean number of singletons per individual, which is the signal measured by SDS (Field et al., 2016). This reduction in overall singleton numbers is consistent with the known effects of population size contraction on reducing tip lengths (Harpending et al., 1998).

With a 10-fold higher mutation rate, there were fewer cases where no individual harbored singletons (Figure 4). Accordingly, SDS scores could be calculated for 65 and 66 out of 100 simulations for the neutral and selective cases, respectively. In these cases, sSDS values were significantly higher in the selected case than for the neutral case (Figure 5; two-sided Wilcox Test $P = 1.1 \times 10^{-5}$). However, note that sSDS values is less than one for the selected case, which does not exceed the FDR threshold in our study (for the high N_0 case, the smallest sSDS value with FDR $<$ 0.05 is 4.46).

Although singleton numbers differ between the two cases, a reduction in power could also be caused by a more general reduction in diversity due to the small recent effective population sizes of cattle. To investigate this effect, we estimated the fixed N_e that would yield the same number of segregating sites in simulations using the standard mutation rate, based on Watterson's estimator (Watterson, 1975; Hudson, 1990; see *Methods* for details). In both cases where selection is present or absent, N_e estimates lie at around 25,000, which is that inferred at approximately halfway between the onset of domestication and the present day (Boitard et al., 2016b; Figure S5). Given that estimates are similar irrespective of whether a sweep was present or not, the reduced population size caused by domestication could have also affected power due to limiting genetic variation and thus the potential to detect subtle sweep signatures associated with polygenic selection.

Overall, these simulations are consistent with population size reductions in *B. taurus* both reducing the overall genetic diversity and the number of singletons, which limits its ability to detect partial sweeps. SDS is more likely to detect signals in regions of elevated mutation rate, suggesting there will likely be an ascertainment bias in where signals are detected in the genome. The reduction in singletons also reduces the power to investigate SDS values in telomeric regions. SDS values are calculated

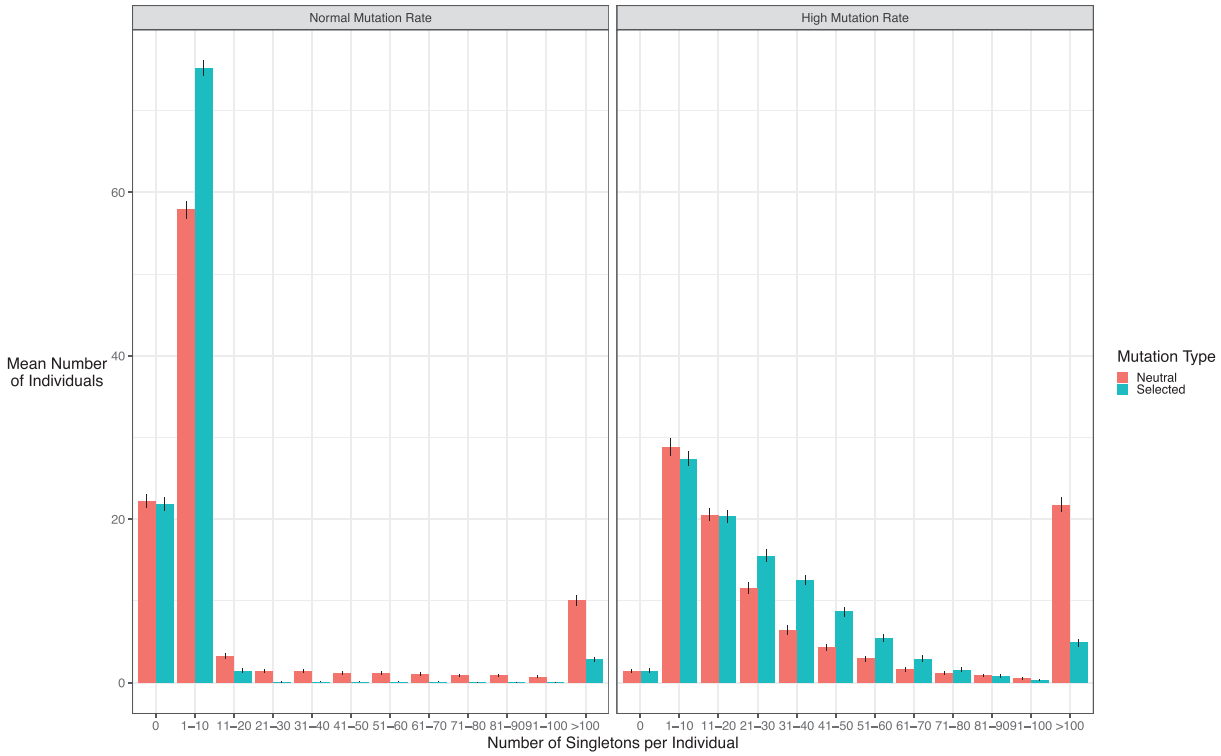

Figure 4. Mean distribution of singleton numbers per individual for each simulation, either assuming a standard mutation rate (left) or a 10-fold higher mutation rate (right). Bars represent 95% confidence intervals.

using the distance up- and downstream from a SNP to the nearest singleton, and are undefined if a certain number of samples do not harbor singletons in either direction (Field et al., 2016). SDS values are hence less likely to be defined in telomeric regions, as it is generally less feasible to observe singletons up until the end of the chromosome. This problem is exacerbated if there are few singletons overall.

Other potential reasons for a lack of signal

Another potential reason for a lack of signal is that the selection response on these traits may have been driven by large-effect variants that have already fixed in the population, with a smaller contribution from small-effect mutations. Theoretical models have shown that more major-effect QTLs are likely to fix if the population lies further from a fitness optimum (Lande, 1983; Jain and Stephan, 2017b; Thornton, 2019). Domesticated species, which experience strong and sustained directional artificial selection, especially in recent generations, could thereby fix more adaptive mutation via sweep-like processes compared to populations evolving in more stable environments (Lande, 1983; Jain & Stephan, 2017a). Furthermore, once a population has adapted to a new environment (the domestication phenotype in this case), then any remaining major-effect mutations are likely to be superseded by variants with weaker effects, which are harder to detect (Hayward & Sella, 2019). The response to polygenic selection will be further weakened in smaller populations (John and Stephan, 2020), which could be a factor given the reduced effective population sizes of *B. taurus* (Sørensen et al., 2005; Boitard et al., 2016b). There is some evidence of this explanation; selective sweeps signatures are associated with stature QTLs (Bouwman et al., 2018), and the study of van den Berg et al. (2020) was more likely to identify milk QTLs that had a moderate to high minor allele frequency, suggesting reduced power to detect low-frequency variants that are potential contributors to polygenic selection. Conversely, the stature meta-analysis by Bouwman et al. (2018) found significant SNPs that explained up to 13.8% of the variance in stature, which is similar to that explained by significant SNPs for human height (16%), which is a classic trait for polygenic selection studies. Hence, there may be sufficient polygenic SNPs present to test for polygenic selection, but the power will still be reduced due to the demographic history of Holstein cattle.

Potential solutions to increase power include increasing sample sizes; using alternative methods; or analyzing different kinds of genome data to detect polygenic selection. Applying SDS to a larger sample size would increase the power to detect selection acting in the recent past (Figure 1; see also Field et al. 2016), but overall power will still be limited by the tip-length of neutral genealogies. Recent developments in methodology involve directly inferring trees from genome data, and using these

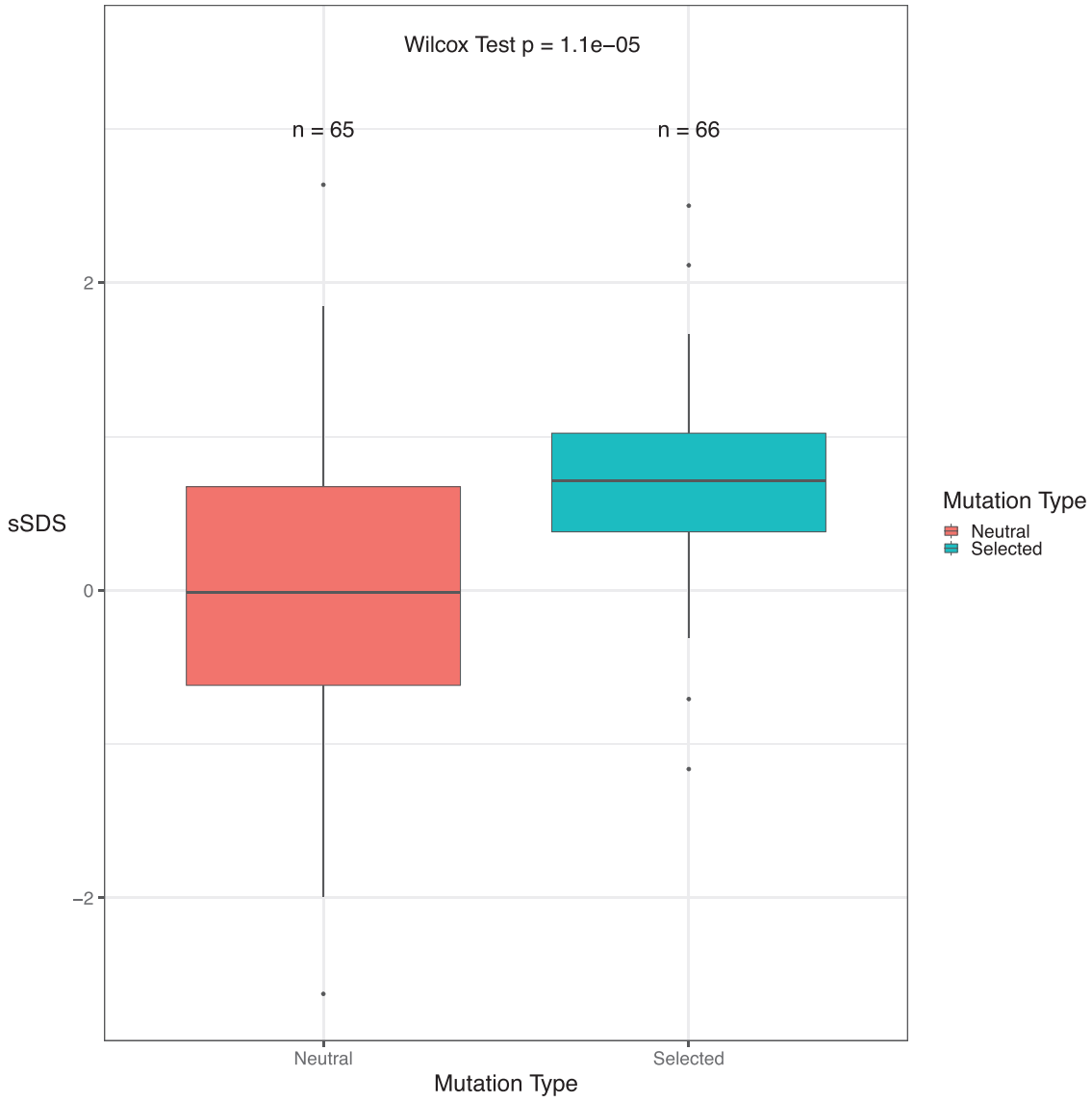

Figure 5. **Distribution of simulated sSDS scores assuming a high mutation rate, for the neutral and selected cases. Numbers above each box plot denote how many simulations produced SDS scores and were included in the plot.**

to identify subtle sweep signatures associated with trait variants (Edge & Coop 2019; Speidel et al., 2019; Stern et al., 2021). These methods have greater power to detect weakly selected mutations that may be segregating for longer than the tip-length of the population.

Another approach would be to look beyond sequence data and focus on gene networks (reviewed by Fagny and Austerlitz 2021). The recently proposed "omnigenic" model (Boyle et al., 2017; Liu et al., 2019) posits that variation in quantitative traits is principally affected by a plethora of "peripheral" genes that indirectly affect them, rather than a limited set of "core" genes that directly modify a trait. These numerous peripheral genes may exert their influence via regulatory effects (e.g., gene expression changes), but are also expected to be highly pleiotropic. Fully testing the omnigenic model will require larger datasets and novel experimental designs (Wray et al., 2018). A recent example is from an experiment with *Drosophila melanogaster*, where gene knockouts that do not pass a GWAS significance threshold for pupal length still significantly affect it (Zhang et al., 2021). There is also nascent evidence that gene regulation may underlie directional polygenic selection. Boitard et al. (2016a) found that some adaptive signatures of *B. taurus* are located in intergenic regions; regulatory changes were also proposed to guide polygenic selection in *Arabidopsis* (He et al., 2016). Analyses of gene sets associated with infection responses or immunity also found evidence for polygenic selection in humans and primates (Daub et al., 2013, 2017; Svardal et al., 2017). Immunity gene sets might be exceptional cases, as they are more likely to contain genes

subject to very strong selection (Castellano et al., 2019). Further investigations using regulatory information and a broader range of gene sets could be a promising approach to determine the impact of polygenic selection.

Materials and Methods

Full methods are available in the Supporting Information.

ACKNOWLEDGMENTS

We would like to thank S. Boitard for sharing his results on *B. taurus* demographic inference and S. Boitard, J. Slate, and two anonymous reviewers for providing feedback on the manuscript. MH is supported by a NERC Independent Research Fellowship (NE/R015686/1). NAP, BG, and TB are funded by a synergistic research grant from the Faculty of Science and Technology, Aarhus University, Denmark. MH and TB also acknowledge financial support from the European Research Council under the European Union's Seventh Framework Program (FP7/20072013, ERC Grant 311341).

AUTHOR CONTRIBUTIONS

All authors contributed to the study design. NAP and BG provided data. MH performed the analyses and wrote the manuscript, with feedback from NAP, BG, and TB.

DATA ARCHIVING

Raw SDS scores and polarization information have been deposited on Dryad (https://doi.org/10.5061/dryad.547d7wm8q). Data analysis and simulation scripts are available on GitHub (https://github.com/MattHartfield/CattleSDS).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

LIERATURE CITED

Barghi, N., J. Hermisson , and C. Schlötterer 2020. Polygenic adaptation: a unifying framework to understand positive selection. Nat. Rev. Genet. 21:769–781.

Barton, N. H., A. M. Etheridge , and A. Véber 2017. The infinitesimal model: definition, derivation, and implications. Theor. Popul. Biol. 118:50–73.

Barton, N. H., J. Hermisson , and M. Nordborg 2019. Why structure matters. eLife 8:e45380.

Barton, N. H., and P. D. Keightley 2002. Understanding quantitative genetic variation. Nat. Rev. Genet. 3:11–21.

van den Berg, I., R. Xiang, J. Jenko, H. Pausch, M. Boussaha, C. Schrooten, et al. 2020. Meta-analysis for milk fat and protein percentage using imputed sequence variant genotypes in 94,321 cattle from eight cattle breeds. Genet. Sel. E 52: 37.

Berg, J. J., and G. Coop 2014. A population genetic signal of polygenic adaptation. PLos Genet. 10:e1004412.

Berg, J. J., A. Harpak, N. Sinnott-Armstrong, A. M. Joergensen, H. Mostafavi, Y. Field, et al. 2019a. Reduced signal for polygenic adaptation of height in UK Biobank. eLife 8:e39725.

Berg, J. J., X. Zhang, and G. Coop 2019b. Polygenic Adaptation has impacted multiple anthropometric traits. bioRxiv https://doi.org/10.1101/167551.

Boitard, S., M. Boussaha, A. Capitan, D. Rocha, and B. Servin 2016a. Uncovering adaptation from sequence data: lessons from genome resequencing of four cattle breeds. Genetics 203: 433–450.

Boitard, S., and D. Rocha 2013. Detection of signatures of selective sweeps in the Blonde d'Aquitaine cattle breed. Anim. Genet. 44:579–583.

Boitard, S., W. Rodríguez, F. Jay, S. Mona, and F. Austerlitz 2016b. Inferring population size history from large samples of genome-wide molecular data - an approximate bayesian computation approach. PLos Genet. 12: e1005877.

Bouwman, A. C., H. D. Daetwyler, A. J. Chamberlain, C. H. Ponce, M. Sargolzaei, F. S. Schenkel, et al. 2018. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. Nat. Genet. 50: 362–367.

Boyle, E. A., Y. I. Li, and J. K. Pritchard 2017. An expanded view of complex traits: from polygenic to omnigenic. Cell 169:1177–1186.

Buitenhuis, B., N. A. Poulsen, G. Gebreyesus, and L. B. Larsen 2016. Estimation of genetic parameters and detection of chromosomal regions affecting the major milk proteins and their post translational modifications in Danish Holstein and Danish Jersey cattle. BMC Genet. 17:114.

Bycroft, C., C. Freeman, D. Petkova, G. Band, L. T. Elliott, K. Sharp, et al. 2018. The UK Biobank resource with deep phenotyping and genomic data. Nature 562:203–209.

Castellano, D., L. H. Uricchio, K. Munch, and D. Enard 2019. Viruses rule over adaptation in conserved human proteins. bioRxiv https://doi.org/10.1101/555060.

Charlesworth, B., and A. W. F. Edwards 2018. A century of variance. Significance 15:20–25.

Chevin, L.-M. 2019. Selective sweep at a QTL in a randomly fluctuating environment. Genetics 213:987–1005.

Chevin, L.-M., and F. Hospital 2008. Selective sweep at a quantitative trait locus in the presence of background genetic variation. Genetics 180:1645–1660.

Coop, G., D. Witonsky, A. Di Rienzo, and J. K. Pritchard 2010. Using environmental correlations to identify loci underlying local adaptation. Genetics 185:1411–1423.

Daub, J. T., T. Hofer, E. Cutivet, I. Dupanloup, L. Quintana-Murci, M. Robinson-Rechavi, et al. 2013. Evidence for polygenic adaptation to pathogens in the human genome. Mol. Biol. E 30: 1544–1558.

Daub, J. T., S. Moretti, I. I. Davydov, L. Excoffier, and M. Robinson-Rechavi 2017. Detection of pathways affected by positive selection in primate lineages ancestral to humans. Mol. Biol. E 34: 1391–1402.

Duntsch, L., B. M. Tomotani, P. deVillemereuil, P. Brekke, K. D. Lee, J. G. Ewen, et al. 2020. Polygenic basis for adaptive morphological variation in a threatened Aotearoa | New Zealand bird, the hihi (*Notiomystis cincta*). Proc. R. Soc. B. 287: 20200948.

Edge, M. D., and G. Coop 2019. Reconstructing the history of polygenic scores using coalescent trees. Genetics 211: 235–262.

Ehrlich, M. A., D. N. Wagner, M. F. Oleksiak, and D. L. Crawford 2021. Polygenic selection within a single generation leads to subtle divergence among ecological niches. Genome Biol. E 13:evaa257.

Exposito-Alonso, M., 500 Genomes Field Experiment Team, H. A. Burbano, O. Bossdorf, R. Nielsen, and D. Weigel 2019. Natural selection on the *Arabidopsis thaliana* genome in present and future climates. Nature, 573:126–129.

Exposito-Alonso, M., F. Vasseur, W. Ding, G. Wang, H. A. Burbano, and D. Weigel 2018. Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. Nat. Ecol. E, 2: 352–358.

Exposito-Alonso, M., P. Wilton, and R. Nielsen. 2020. Non-additive polygenic models improve predictions of fitness traits in three eukaryote model species. bioRxiv https://doi.org/10.1101/2020.07.14.194407.

Fagny, M., and F. Austerlitz 2021. Polygenic adaptation: integrating population genetics and gene regulatory networks. Trends Genet. 37:631–638.

Field, Y., E. A. Boyle, N. Telis, Z. Gao, K. J. Gaulton, D. Golan, et al. 2016. Detection of human adaptation during the past 2000 years. Science 354:760–764.

Fisher, R. A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 52:399–433.

Fuhrmann, N., C. Prakash, and T. S. Kaiser 2021. Polygenic adaptation from standing genetic variation allows rapid ecotype formation. bioRxiv https://doi.org/10.1101/2021.04.16.440113.

Gazal, S., H. K. Finucane, N. A. Furlotte, P.-R. Loh , P. F. Palamara, X. Liu, et al. 2017. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nat. Genet. 49:1421–1427.

Georges, M., C. Charlier, and B. Hayes 2019. Harnessing genomic information for livestock improvement. Nat. Rev. Genet. 20:135–156

Gramlich, S., X. Liu, A. Favre, C. A. Buerkle, and S. Karrenberg 2021. A polygenic architecture with conditionally neutral effects underlies ecological differentiation in *Silene*. bioRxiv https://doi.org/10.1101/2021.07.06.451304.

Haile-Mariam, M., G. J. Nieuwhof, K. T. Beard, K. V. Konstatinov, and B. J. Hayes 2013. Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations: implication of genomic heritability for genomic evaluation. J. Anim. Breed. Genet. 130:20–31.

Harland, C., W. Coppieters, E. Mullaart, C. Charlier, and M. Georges 2018. Rate of de novo mutation in dairy cattle and potential impact of reproductive technologies. *In* Proceedings of the World Congress on Genetics Applied to Livestock Production.

Harpending, H. C., M. A. Batzer, M. Gurven, L. B. Jorde, A. R. Rogers, and S. T. Sherry 1998. Genetic traces of ancient demography. Proc. Natl. Acad. Sci. USA 95:1961–1967.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard 2009. Invited review: genomic selection in dairy cattle: progress and challenges. J. Dairy Sci. 92:433–443.

Hayward, L. K., and G. Sella 2019. Polygenic adaptation after a sudden change in environment. bioRxiv https://doi.org/10.1101/792952.

He, F., A. L. Arce, G. Schmitz, M. Koornneef, P. Novikova, A. Beyer, et al. 2016. The footprint of polygenic adaptation on stress-responsive cis-regulatory divergence in the *Arabidopsis* genus. Mol. Biol. E 33:2088–2101.

Hudson, R. R. 1990. Gene genealogies and the coalescent process. In Oxford surveys in evolutionary biology (ed. by Futuyma, D.J. & Antonovics, J.). Oxford Univ. Press, Oxford, U.K., pp. 1–42.

Jain, K., and W. Stephan 2015. Response of polygenic traits under stabilizing selection and mutation when loci have unequal effects. G3 5:1065–1074.

Jain, K., and W. Stephan 2017a. Modes of rapid polygenic adaptation. Mol. Biol. E 34:3169–3175.

Jain, K., and W. Stephan 2017b. Rapid adaptation of a polygenic trait after a sudden environmental shift. Genetics 206:389–406.

Jiménez-Mena, B., P. Tataru, R. F. Brøndum, G. Sahana, B. Guldbrandtsen, and T. Bataillon 2016. One size fits all? Direct evidence for the heterogeneity of genetic drift throughout the genome. Biol. Lett. 12:20160426.

John, S., and W. Stephan 2020. Important role of genetic drift in rapid polygenic adaptation. Ecol. E 10:1278–1287.

Josephs, E. B., J. J. Berg, J. Ross-Ibarra, and G. Coop 2019. Detecting adaptive differentiation in structured populations with genomic data and common gardens. Genetics 211:989–1004.

Kreiner, J. M., P. J. Tranel, D. Weigel, J. R. Stinchcombe, and S. I. Wright 2020. The genetic architecture and genomic context of glyphosate resistance in *Amaranthus tuberculatus*. bioRxiv https://doi.org/10.1101/2020.08.19.257972.

Lande, R. 1983. The response to selection on major and minor mutations affecting a metrical trait. Heredity 50:47–65.

Le Corre, V., and A. Kremer 2012. The genetic differentiation at quantitative trait loci under local adaptation. Mol. Ecol. 21:1548–1566.

Lemay, D. G., D. J. Lynn, W. F. Martin, M. C. Neville, T. M. Casey, G. Rincon, et al. 2009. The bovine lactation genome: insights into the evolution of mammalian milk. Genome Biol. 10:R43.

Liu, X., Y. I. Li , and J. K. Pritchard 2019. Trans effects on gene expression can drive omnigenic inheritance. Cell 177:1022–1034.e6.

MacEachern, S., B. Hayes, J. McEwan, and M. Goddard 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC Genom 10:181.

MacLachlan, I. R., T. K. McDonald, B. M. Lind, L. H. Rieseberg, S. Yeaman, and S. N. Aitken 2021. Genome-wide shifts in climate-related variation underpin responses to selective breeding in a widespread conifer. Proc. Natl. Acad. Sci. USA 118:e2016900118.

Meuwissen, T., B. Hayes, and M. Goddard 2013. Accelerating improvement of livestock with genomic selection. Annu. Rev. Anim. Biosci. 1:221–237.

Nielsen, R. 2005. Molecular signals of natural selection. Annu. Rev. Genet. 39:197–218.

Novembre, J., and N. H. Barton 2018. Tread lightly interpreting polygenic tests of selection. Genetics 208:1351–1355.

Pavlidis, P., D. Metzler, and W. Stephan 2012. Selective sweeps in multilocus models of quantitative traits. Genetics 192:225–239.

Pritchard, J. K., and A. Di Rienzo 2010. Adaptation - not by sweeps alone. Nat. Rev. Genet. 11:665–667.

Pritchard, J. K., J. K. Pickrell , and G. Coop 2010. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. Curr. Biol. 20: R208–R215.

Qanbari, S., H. Pausch, S. Jansen, M. Somel, T. M. Strom, R. Fries, et al. 2014. Classic selective sweeps revealed by massive sequencing in cattle. PLoS Genet. 10: e1004148.

Qanbari, S., E. C. G. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, et al. 2010. A genome-wide scan for signatures of recent selection in Holstein cattle. Anim. Genet. 41:377–389.

Racimo, F., J. J. Berg, and J. K. Pickrell 2018. Detecting polygenic adaptation in admixture graphs. Genetics 208:1565–1584.

Robinson, M. R., G. Hemani, C. Medina-Gomez, M. Mezzavilla, T. Esko, K. Shakhbazov, et al. 2015. Population genetic differentiation of height and body mass index across Europe. Nat. Genet. 47:1357.

Rowan, T. N., H. J. Durbin, C. M. Seabury, R. D. Schnabel, and J. E. Decker 2021. Powerful detection of polygenic selection and evidence of environmental adaptation in US beef cattle. PLos Genet. 17:e1009652.

Sanjak, J. S., J. Sidorenko, M. R. Robinson, K. R. Thornton, and P. M. Visscher 2018. Evidence of directional and stabilizing selection in contemporary humans. Proc. Natl. Acad. Sci. USA 115:151–156.

Savolainen, O., M. Lascoux, and J. Merila 2013. Ecological genomics of local adaptation. Nat. Rev. Genet. 14:807–820.

Schoech, A. P., D. M. Jordan, P. R. Loh, S. Gazal, L. J. O'Connor, D. J. Balick, et al. 2019. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. Nat. Commun. 10:790.

Sella, G., and N. H. Barton 2019. Thinking about the evolution of complex traits in the era of genome-wide association studies. Annu. Rev. Genom. Hum. Genet. 20:461–493.

Sohail, M., R. M. Maier, A. Ganna, A. Bloemendal, A. R. Martin, M. C. Turchin, et al. 2019. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife 8:e39702.

Sørensen, A. C., M. K. Sørensen, and P. Berg 2005. Inbreeding in danish dairy cattle breeds. J. Dairy Sci. 88:1865–1872.

Soyeurt, H., A. Gillon, S. Vanderick, P. Mayeres, C. Bertozzi, and N. Gengler 2007. Estimation of heritability and genetic correlations for the major fatty acids in bovine milk. J. Dairy Sci. 90:4435–4442.

Speidel, L., M. Forest, S. Shi, and S. R. Myers 2019. A method for genome-wide genealogy estimation for thousands of samples. Nat. Genet. 51:1321–1329.

Stephan, W. 2016. Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. Mol. Ecol. 25:79–88.

Stephan, W. 2019. Selective sweeps. Genetics 211:5–13.

Stern, A. J., L. Speidel, N. A. Zaitlen, and R. Nielsen 2021. Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. Am. J. Hum. Genet. 108:219–239.

Stetter, M. G., K. Thornton, and J. Ross-Ibarra 2018. Genetic architecture and selective sweeps after polygenic adaptation to distant trait optima. PLos Genet. 14: e1007794.

Svardal, H., A. J. Jasinska, C. Apetrei, G. Coppola, Y. Huang, C. A. Schmitt, et al. 2017. Ancient hybridization and strong adaptation to viruses across African vervet monkey populations. Nat. Genet. 49:1705–1713.

Thornton, K. R. 2019. Polygenic adaptation to an environmental shift: temporal dynamics of variation under Gaussian stabilizing selection and additive effects on a single trait. Genetics 213:1513–1530.

Turchin, M. C., C. W. Chiang, C. D. Palmer, S. Sankararaman, D. Reich, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, et al. 2012. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nat. Genet. 44:1015–1019.

Uricchio, L. H., H. C. Kitano, A. Gusev, and N. A. Zaitlen 2019. An evolutionary compass for detecting signals of polygenic selection and mutational bias. Evol. Lett. 3: 69–79.

Visscher, P. M., and M. E. Goddard 2019. From R.A. Fisher's 1918 paper to GWAS a century later. Genetics 211:1125–1130.

Vitti, J. J., S. R. Grossman , and P. C. Sabeti 2013. Detecting natural selection in genomic data. Annu. Rev. Genet. 47:97–120.

Vladar, H. P., and N. Barton 2014. Stability and response of polygenic traits to stabilizing selection and mutation. Genetics 197:749–767.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. Theor. Popul. Biol. 7:256–276.

Whitlock, M. C. 2008. Evolutionary inference from F_{ST}. Mol. Ecol. 17:1885–1896.

Wieters, B., K. A. Steige, F. He, E. M. Koch, S. E. Ramos-Onsins, H. Gu, et al. 2021. Polygenic adaptation of rosette growth in *Arabidopsis thaliana*. PLoS Genet. 17: e1008748.

Wollstein, A., and W. Stephan 2014. Adaptive fixation in two-locus models of stabilizing selection and genetic drift. Genetics 198:685–697.

Wray, N. R., K. E. Kemper, B. J. Hayes, M. E. Goddard, and P. M. Visscher 2019. Complex trait prediction from genome data: contrasting EBV in livestock to PRS in humans. Genetics 211:1131–1141.

Wray, N. R., C. Wijmenga, P. F. Sullivan, J. Yang and P. M. Visscher 2018. Common disease is more complex than implied by the core gene omnigenic model. Cell 173:1573–1580.

Yang, J., A. Bakshi, Z. Zhu, G. Hemani, A. A. E. Vinkhuyzen, S. H. Lee, et al. 2015. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat. Genet. 47:1114–1120.

Yeaman, S., K. A. Hodgins, K. E. Lotterhos, H. Suren, S. Nadeau, J. C. Degner, et al. 2016. Convergent local adaptation to climate in distantly related conifers. Science 353: 1431.

Zan, Y., and Ö. Carlborg 2018. A polygenic genetic architecture of flowering time in the worldwide *Arabidopsis thaliana* population. Mol. Biol. E 36:141–154.

Zeder, M. A. 2008. Domestication and early agriculture in the Mediterranean Basin: origins, diffusion, and impact. Proc. Natl. Acad. Sci. USA 105:11597–11604.

Zeng, J., R. deVlaming, Y. Wu, M. R. Robinson, L. R. Lloyd-Jones, L. Yengo, et al. 2018. Signatures of negative selection in the genetic architecture of human complex traits. Nat. Genet. 50: 746–753.

Zeng, J., A. Xue, L. Jiang, L. R. Lloyd-Jones, Y. Wu, H. Wang, et al. 2021. Widespread signatures of natural selection across human complex traits and functional genomic categories. Nat. Commun. 12:1164.

Zhang, W., G. R. Reeves, and D. Tautz 2021. Testing implications of the omnigenic model for the genetic analysis of loci identified through genome-wide association. Curr. Biol. 31:1092–1098.e6.

Zhao, F., S. McParland, F. Kearney, L. Du, and D. P. Berry 2015. Detection of selection signatures in dairy and beef cattle using high-density genomic information. Genet. Sel. E 47:49.

Zhou, X., P. Carbonetto, and M. Stephens 2013. Polygenic modeling with bayesian sparse linear mixed models. PLoS Genet. 9:e1003264.

Associate Editor: Z. Gompert

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1: Genome wide sSDS results for the low N_0 model.
Figure S2: Polygenic selection test results for the high N_0 model (milk fat percentage, milk protein percentage) after removing outlier point.
Figure S3: Polygenic selection test results for the low N_0 model.
Figure S4: Permutation test results.
Figure S5: N_e estimates from simulations.
Figure S6: Histogram of mean *F*-values for each individual.
Figure S7: Schematic of data filtering.
Figure S8: Density of mean site depth.
Figure S9: Example of sliding-window analysis to detect over-assembled regions (OARs).
Table S1: List of putative OARs.
Table S2: Cutoff values for filtering singletons.